

APPLICATION FOR UNITED STATES PATENT

in the name of

Lorin Sutton, Craig Despaux, and Mike Adamski

of

America Online, Inc.

for

Identifying Unwanted Electronic Messages

Fish & Richardson P.C.
601 Thirteenth Street, NW
Washington, DC 20005
Tel.: (202) 783-5070
Fax: (202) 783-2331

ATTORNEY DOCKET:

06975-211001

Identifying Unwanted Electronic Messages

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/286,963 filed April 30, 2001, which is incorporated by reference.

TECHNICAL FIELD

5 This invention relates to the identification of unwanted electronic messages in a message exchanging system.

BACKGROUND

10 Through the exchange of electronic messages, a new medium of communication has evolved. As this new communication medium has become more pervasive, growth has been experienced both in the electronic networks supporting electronic messages and the number of people having access to those electronic networks. With this growth, message exchangers have been subject to an increasing number of spam and other unwanted messages, as well as hacker attacks through electronic messaging.

SUMMARY

15 In one general aspect, the performance of a message exchanging system may be improved. A payload portion of a message being communicated is inspected and characteristics of the payload portion are identified and compared with stored data indicating characteristics of at least one other message that has been inspected. A security condition is identified based on the comparison.

20 In another general aspect, the performance of a message exchanging system may be improved by inspecting a message being communicated to a first device in a message exchanging system of two or more devices and identifying characteristics of the message. Characteristics of the message are compared with stored data indicating characteristics of at least one other message communicated to a second device, and a security condition is
25 identified based on the comparison.

Implementations may include one or more of the following features. For example, the characteristics of the payload portion include information other than address information. The characteristics of the payload portion inspected do not include address information. The message exchanged may include an electronic mail message.

5 The characteristics may be tracked for comparison against characteristics of future messages, and the characteristics of a new message may be compared with the characteristics of at least one message that has been tracked. Comparing characteristics may include comparing characteristics with stored characteristics of other communicated messages.

10 Implementations may include rejecting the message if the security condition identified includes a hostile indicator. The hostile indicator may be revealed as a hostile indicator when comparing characteristics of the messages inspected reveals a threshold of messages having a shared characteristic.

15 The security condition may include an indeterminate indicator. Implementations may include determining that the security condition includes an indeterminate indicator when the characteristics, standing alone, do not reveal a hostile security condition, but the characteristics may do so in combination with similar characteristics of other messages, including those exchanged in the future. Implementations may include removing messages with these characteristics if these characteristics subsequently generate a hostile indicator for a security condition. The message may be accepted if the security condition includes an
20 indeterminate indicator.

 Implementations may include generating a neutral indicator for the security condition. If the security condition includes a neutral indicator, the message exchanging system may accept the message.

25 Implementations also may include inspecting messages sent or received by more than a single device.

 Implementations may include a system capable of achieving the above features, for instance, a remote exchanging system, a local exchanging system, and a network between these components. Implementations also may include rearranging the sequence of steps performed on the local exchanging system to achieve these features.

The details of one or more implementations are set forth in the accompanying drawings and the description below. Other features, and advantages will be apparent from the description and drawings.

DESCRIPTION OF DRAWINGS

Fig. 1 is a diagram of a message exchanging system with the ability to examine exchanged messages for unwanted messages.

Fig. 2 is a diagram of an exemplary structure of message that may be exchanged in a communications system such as that shown in Fig. 1.

Figs. 3-6 are flow charts illustrating steps performed in exchanging a message.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

For illustrative purposes, Figs. 1-6 describe message exchanging systems and processes capable of determining whether an electronic message being communicated is unwanted. Generally, a message exchanging system inspects an exchanged message by determining one or more characteristics of the message and comparing them to one or more characteristics found in other messages. Based on this comparison, a security condition may be identified, and a responsive action taken. For instance, the message may be discarded if the security condition is deemed hostile. The message may be accepted if the security condition is deemed neutral or better, or the message may be tracked if the security condition is deemed indeterminate to enable responsive action based on future or other comparisons involving the characteristics of this or another exchanged message.

Referring to Fig. 1, a message exchanging system 100 may be structured and arranged to transmit messages between a remote exchanging system 110 and a local exchanging system 130 through a network 120. For brevity, each of these elements is represented as a monolithic entity. However, any or all of system 110, the network 120 and the system 130 may include numerous interconnected computers and components designed to perform a set of specified operations and/or dedicated to a particular geographical region.

Typically, the remote exchanging system 110 and the local exchanging system 130 are structured and arranged to exchange one or more messages across network 120. Each of the remote exchanging system 110 and the local exchanging system 130 may be implemented

by a general-purpose computer capable of responding to and executing instructions in a defined manner. Each of the remote exchanging system 110 and the local exchanging system 130 may include a personal computer, a special-purpose computer, a workstation, a server, a device, a component, other equipment or some combination thereof capable of responding to and executing instructions. Each may be structured and arranged to receive instructions from, for example, a software application, a program, a piece of code, a device, a computer, a computer system, or a combination thereof, which independently or collectively directs operations, as described herein. The instructions may be embodied permanently or temporarily in any type of machine, component, equipment, storage medium, or propagated signal that is capable of being delivered to the remote exchanging system 110 or the local exchanging system 130.

One example of the remote exchanging system 110 includes a dedicated mailing system. Such a dedicated mailing system may be implemented by specialized hardware or executed by a general purpose processor capable of running various applications such as electronic mailer programs, either or both being capable of employing various message transfer protocols such as SMTP ("Simple Mail Transfer Protocol"). In addition or as an alternative, the remote exchanging system 110 may include a communications interface (not shown) in an information delivery network. For example, the remote exchanging system 110 may include an electronic mail gateway.

In any event, the remote exchanging system 110 generally communicates with the local exchanging system 130 using network 120. As such, the network 120 typically is structured and arranged to enable direct or indirect communications between the remote exchanging system 110 and the local exchanging system 130.

Examples of the network 120 include the Internet, the World Wide Web, one or more WANs ("Wide Area Networks"), one or more LANs ("Local Area Networks"), one or more analog or digital wired or wireless telephone networks (e.g., PSTN ("Public Switched Telephone Network"), ISDN ("Integrated Services Digital Network"), or xDSL ("Digital Subscriber Loop") network), a radio, a television, a cable, a satellite, and/or other delivery mechanisms for carrying data. The network 120 may include a direct link between the remote exchanging system 110 and the local exchanging system 130, or the network 120 may include one or more networks or subnetworks between them. Each network or subnetwork

may include, for example, a wired or wireless data pathway capable of carrying and receiving data between remote exchanging system 110 and local exchanging system 130.

Typically, the local exchanging system 130 is structured and arranged to exchange one or more messages with remote exchanging system 110 across network 120. The local exchanging system 130 may include or form part of an information delivery system, such as, for example, an electronic mail system, the World Wide Web, or an online service provider network. The local exchanging system 130 is structured and arranged to receive one or more messages.

The local exchanging system 130 may include various components, including one or more of an inspection module 131, a comparison module 132, a data store of characteristics 133, and a security module 134, as illustrated by Fig. 1. In general, each of the modules and data store 133 may be independently or collectively implemented by, for example, a general-purpose computer.

The inspection module 131 may be structured and arranged to exchange and analyze a message or one or more characteristics of the message or its payload portion when communicated with one or more devices, such as another local exchanging system 130A.

The comparison module 132 may be structured and arranged to compare the characteristics of the payload portion of the inspected message with a data store of characteristics 133 or to compare the characteristics of messages exchanged across more than one device with a data store of characteristics 133.

The data store 133 may be structured and arranged to include a compilation of suspect message characteristics identified as potentially problematic, suspicious or profile-matching. Examples of such characteristics include, but are not limited to, the existence or attributes of text, a key word, a name, a physical size and/or content of an attached file, and the address of hyper text embedded in a message. When messages exchanged across more than one device are inspected, and collectively used to identify unwanted or suspect messages, characteristics stored in data store 133 also may include an IP address, a sender identification and domain name information (e.g., *name.com*).

Implementations of the data store 133 may include database software structured and arranged to manage information relating to characteristics of the messages. For example, the database software may keep a table of entries or terms that the local exchanging system 130

is inspecting and tracking. Each entry may include a counter indicating the number of times the entry has appeared. The entry also may include a location parameter including addresses or message identifiers indicating messages in which the entry appears. Referencing this location parameter enables retrieval of messages subsequently determined to be unwanted.

5 The security module 134 may be structured and arranged to identify a security condition based on results from the comparison module 132.

Although described above with respect to a single local exchanging system 130, the message exchanging system 100 may include more than one local exchanging system 130 structured and arranged to communicate messages, as depicted by local exchanging system 10
130A in Fig. 1. For example, an organization may use multiple servers capable of exchanging messages and may distribute messages to be communicated across the multiple servers in a manner that balances the load.

Fig. 2 shows an exemplary structure of a message 200 of the type exchanged in Fig. 1. In general, the message 200 may include, e.g., an electronic mail message and a file attachment. The message 200 may be structured and arranged to include a header field 210 and a payload portion 220. The header field 210 typically includes addressing information to describe the destination of the message. The header field 210 may include an IP address, a mail recipient identifier, a PC identifier, and/or an online identity. The payload portion 220 typically includes information other than address or identification information, such as
15 information to be communicated to the person or system identified by the header field 210. For instance, the payload portion field 220 may include a letter in an electronic mail message, an attached file in an electronic mail message, or a hypertext link in a file.

Fig. 3 illustrates a method of identifying unwanted messages in a message exchanging system, such as local exchanging system 130 described with reference to Fig. 1. Typically, an unwanted message may be identified by identifying a security condition for a
25 message (step 310), determining whether the security condition is hostile, indeterminate or neutral (step 320), and taking an action based on the security condition identified (steps 330, 340 and 350).

Initially, a local exchanging system identifies a security condition for a message
30 exchanged (step 310). Implementations may include systems that inspect the payload portion

of a message. An example of operations performed by systems that inspect the payload portion are described further with respect to Fig. 4.

Implementations also may include systems that inspect both the header field and the payload portion. Such systems may be used where messages are exchanged across more than one local exchanging system.

Generally, identifying a security condition involves comparing one or more parameters appearing in a message with stored data indicating that the message may be hostile. The stored data generally indicate characteristics of at least one other message previously inspected.

The local exchanging system then determines whether the security condition is hostile, neutral, or indeterminate (step 320). A hostile security condition indicates that, based on parameters of the message, the message has a profile that resembles an unwanted message (e.g., spam, objectionable content) or a malicious message (e.g., viruses, worms).

A neutral security condition indicates that, based on the parameters of the message and based on the data presently stored, the message does not resemble messages considered to be unwanted or malicious.

An indeterminate condition indicates that, based on the parameters of the message, the message has a profile that is of concern and may subsequently be identified as a hostile message. For example, an exchanging system may receive a large number of messages from one source. After a threshold number of messages are exchanged, the message may be identified as a hostile message. Messages leading to the threshold number may initially generate a neutral, then an indeterminate indicator, before the threshold iteration of the message generates a hostile indicator.

If the message is hostile, the local exchanging system rejects the message (step 330). In the case of a message being transmitted, rejecting the message may include not transmitting the message. In the case of messages being received, storage and processing of rejected messages may be prevented, or to the extent that rejected messages are stored, an alarm may be generated and/or sent to an administrator.

The local exchanging system generally processes (e.g., transmit or receive) messages for which the security condition includes a neutral indicator indicating that the characteristics of the exchanged message correspond to those messages considered not hostile (step 340).

The local exchanging system also generally processes messages for which the security condition includes an indeterminate indicator, as this security condition indicates that the characteristics of the message do not correspond to a hostile condition at this time but may reveal a hostile indicator in the future in combination with other received messages having similar characteristics (step 350). As part of processing a message with an indeterminate indicator, the local exchanging system may index the message that has been processed to enable subsequent action to be taken if the message is recategorized. Similarly, characteristics may be counted to better categorize the message.

Fig. 4 illustrates a procedure 400 that represents one method of identifying a security condition by inspecting the payload portion of a message in a message exchanging system. Procedure 400 includes exchanging a message (step 410), inspecting the payload portion of the message (step 420), comparing the characteristics of the payload portion of the message with a data store of characteristics of other messages (step 430), and identifying a security condition based on the comparison of the characteristics (step 440). Typically, procedure 400 is performed by a message exchanging system, such as local exchanging system 130 of Fig. 1.

A message or file is exchanged between a sender and a receiver, such as remote exchanging system 110 and local exchanging system 130 (step 410). The message may include an electronic mail message and/or an instant message, and the message may be transmitted to or from a local exchanging system.

Next, in the implementation of Fig. 1, the inspection module 131 inspects the payload portion of the message exchanged (step 420). The payload portion generally corresponds to the payload portion of the message described previously in Fig 2. Implementations may include inspecting more than one field in the payload portion. For example, the local exchanging system 130 may inspect the exchanged message to determine if the message includes hypertext links and/or attached documents. If the message includes a reference to information located outside the message, the external information being referenced also may be inspected. For example, in a message with a link to a file on a server, the local exchanging system may download and inspect the file.

The comparison module 132 compares the payload portion of the exchanged message, or characteristics thereof, with information from a data store 133 (step 430). This

information may include the payload portion, or characteristics thereof, of other exchanged messages that have been inspected. Where the data store includes a database of the characteristics, the local exchanging system 130 may compare characteristics of the payload portion to those of other messages and add the compared characteristics to the data store.

5 The data store then may be updated as additional messages are received. Other implementations may include having an administrator set parameters to inspect. For example, if an administrator learns in advance of a virus, the administrator may specify that all files with a suspect name or profile be entered into the data store of characteristics.

10 Comparing the characteristics may include comparing characteristics of an exchanged message with a subset of characteristics of other messages. For example, a local exchanging system may filter characteristics in the data store so that characteristics of an exchanged message are compared against the filtered subset of more suspect characteristics. The characteristics of the exchanged message may still be compiled into the data store. These characteristics may “bubble” into the filtered characteristics that are compared against if the characteristics continue to be received or are recategorized as more suspect.

15 In another implementation, the message may be compared against a data store corresponding to characteristics for messages exchanged locally. For example, a data store may correspond to messages exchanged on that system in a specified time span.

20 In the implementation of Fig. 1, the security module 134 determines a security condition based on the results of the comparison of the payload portion in the comparison module 132 (step 440).

25 Fig. 5 illustrates a procedure 500 by which a security condition is identified by inspecting both the header field and the payload portion of messages being communicated in a local exchanging system that includes two or more devices. Procedure 500 involves exchanging a message (step 510), inspecting the message (step 520), comparing characteristics of the message with a data store of characteristics of other messages (step 530), and identifying a security condition based on the comparison (step 540). Typically, procedure 500 is performed on a message exchanging system, such as that illustrated by local exchanging system 130 of Fig. 1.

30 Initially, a message is exchanged (step 510) between a remote exchanging system and a local exchanging system, as is depicted in dashed lines in Fig. 1.

5 The exchanged message then is inspected (step 520) by examining parameters both in the header field and the payload portion. The local exchanging system then compares the message inspected with characteristics of messages exchanged across more than one local exchanging system 130 (step 530). The message exchanged across more than one local exchanging system 130 may be acquired in a synchronous or disparate manner. For example, the characteristics of messages compared may include characteristics of messages compiled from one server sending messages and another server receiving messages. The two or more local exchanging servers may be situated in geographically diverse locations. For example, one local exchanging server may be located on the east coast while the other is located on the west coast.

10 Comparing characteristics of messages (step 530) may include using a counter in conjunction with characteristics to determine a security condition. For example, a database may keep track of the number of times certain characteristics appear. As will be discussed, the counter may be a factor in determining the security condition.

15 The local exchanging system then identifies a security condition (step 540) based on the result of the comparison with messages exchanged across more than one local exchanging system. The security condition may include a hostile indicator.

20 Determining that there is a hostile indicator may include tracking the number of suspect elements in a message. A characteristic of the message is a suspect element when that characteristic is identified in the comparison against entries in the data store of characteristics 133. For example, if a Uniform Resource Locator ("URL") found in a message also exists in the data store of characteristics 133, that correlation may be identified as a suspect element that implicates the message as a suspect message.

25 Determining that there is a hostile indicator may include quantifying suspect elements. For example, two suspect elements may generate an indeterminate indicator while three suspect elements generate a hostile indicator.

30 In addition, or as an alternative, the security condition may be identified depending on the actual suspect element detected within the message. For example, messages with one particular suspect element H and no other elements of concern may always generate a hostile indicator while messages with a different single suspect element or a combination of other suspect elements may not generate a hostile indicator. Likewise, a message may include five

suspect elements, but if one of the elements is a particular suspect element, the message may generate a neutral indicator. Examples may feature a hierarchy of suspect elements where one particular suspect element generates a neutral indicator unless another suspect element is present, in which case a hostile indicator is generated.

5 Implementations also may include having a suspect element generate an alarm score to gauge the level of concern. For example, a message may be inspected by identifying a sender, an attached file and a MD5 ("Message Digest 5") signature as elements of concern. The sender may receive a score of 10, the attached file may receive a score of 20, and the MD5 signature may generate a score of 30 for a combined message score of 60. If the local
10 exchanging system categorizes all messages with a score greater than 100 as hostile, the message may be considered indeterminate or neutral. However, in some implementations, if one of the elements of concern is exchanged with increasing frequency, the score associated with that element of concern may increase. Thus, if the sender continues to appear in
15 messages exchanged, perhaps indicating the sender may be sending "spam" mail messages, the score associated with that sender may rise to 90, generating a new alarm score of 140 for the same message previously assigned a score of 60. In some implementations, messages having alarm scores that subsequently increase above a specified threshold may be deleted in response to such an increase. For example, a local exchanging system may categorize a
20 message as hostile initially if the score is above 100 and subsequently re-categorize as hostile any messages whose score rises above 130. In this case, the message is categorized with a hostile indicator upon review and the message is deleted. The local exchanging system may look up messages that were initially categorized with indeterminate indicators and subsequently re-categorized as hostile, and delete the re-categorized messages.

25 Determining a security condition also may include using neural networks to categorize and classify messages. The use of neural networks enables a local exchanging system to "learn" based on changing message patterns and conditions.

Implementations also may include tracking messages that include an indeterminate indicator. Generally, these implementations apply to situations where the local exchanging system has permissions over other systems, but are not limited to such situations.

30 Implementations in which the remote system is operated by a different entity may employ a protocol to allow the tracking of messages between the entities. For example, messages A, B

and C each include characteristic Z, which may generate a hostile indicator if the characteristic Z occurs above a threshold number of times. The local exchanging system 130 may store messages A, B, and C, but will track the addresses at which the messages are located. If the local exchanging system 130 exchanges message D with characteristic Z, and the threshold number of times for characteristic Z to generate a hostile indicator is four or more times, then the local exchanging system 130 may reject message D. The local exchanging system 130 also may delete messages A, B, and C in response to the threshold having been reached, even after initially processing them.

Fig. 6 illustrates a procedure 600 by which a message with an indeterminate indicator is tracked, as was described generally in step 380 of Fig. 3. The implementations used to identify the security condition may include, but are not limited to, the steps described with respect to Figs. 4 and 5. Procedure 600 is typically performed on a message exchanging system, such as local exchanging system 130 of Fig. 1.

Initially, an exchanged message with an indeterminate indicator is processed (step 610). This generally includes transmitting a message or storing a received message.

With the message processed, the local exchanging system tracks the location of where the message is kept (step 620). Typically, this will include having a message exchanging system track the location of a message. However, implementations may include having the local exchanging system receive a location of the message from a remote exchanging system indicating where the message is kept. Other implementations of tracking the message may include tracking an instance of the message being stored in an "outbox" of sent messages on a local exchanging system. The location of this message also may be provided.

The local exchanging system exchanges additional messages (step 630). As these additional messages are exchanged, the additional messages are inspected (step 640). Inspecting the additional messages includes examining the subsequently received messages to determine whether they are unwanted (e.g., Figs. 3-5) and updating the data store of characteristics.

The local exchanging system determines whether updating the data store of characteristics with characteristics of messages subsequently exchanged recategorizes a message previously categorized with an indeterminate indicator into a message with a hostile indicator (step 650). If so, the message is removed from storage (step 660). In cases where

the message was transmitted, the local exchanging system may generate a message, alarm or indicator to the remote exchanging system that the message is now considered to have a hostile indicator. If the message has not been recategorized, the local exchanging system continues to track the message (step 670).

5 Implementations also may include recategorizing messages with indeterminate indicators into neutral indicators if subsequently exchanged messages indicate that the message is valid. For example, a valid message sender may send valid electronic mail to a large number of recipients, such that the number of recipients happens to be more than the threshold required to generate an indeterminate indicator. In another example, a system
10 administrator who receives an alarm about a particular profile in a message may examine the message and determine that the message is acceptable to be stored.

The message exchanging system, methods, devices and programs may be implemented in hardware or software, or a combination of both. In some implementations, the message exchanging system, methods, devices and programs are implemented in
15 computer programs executing on programmable computers each with at least one processor, a data storage system (including volatile and/or storage elements), at least one input device, and at least one output device. Program code is applied to input data to perform the functions described herein and generate output information. The output information is applied to one or more output devices.

20 A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made.